DOCUMENT RESUME

ED 285 914 TM 870 522

AUTHOR Fabrey, Lawrence J.; Raymond, Mark R.
TITLE Congruence of Standard Setting Methods for a Nursing Certification Examination.
PUB DATE Apr 87
NOTE 7p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Washington, DC, April 21-23, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adults; *Certification; *Cutting Scores; Judgment Analysis Technique; *Licensing Examinations (Professions); Mastery Tests; *Nursing; Peer Evaluation; Professional Recognition; Standards; *Test Interpretation
IDENTIFIERS American Nurses Association

ABSTRACT

The American Nurses' Association certification provides professional recognition beyond licensure to nurses who pass an examination. To determine the passing score as it would be set by a representative peer group, a survey was mailed to a random sample of 200 recently certified nurses. Three questions were asked: (1) what percentage of examinees should pass; (2) what percentage of questions should be answered correctly for certification; and (3) what score should be achieved for certification, given a hypothetical distribution of scores on a 75-item test. There were 98 responses. Respondents indicated that a mean of 70.36% should pass. The currently used Angoff technique resulted in a passing rate of 87.96%. The mean percentage of questions that should be answered correctly was calculated to be 71.28% according to the survey, 56% as set by committee. The mean score that should be achieved for certification was calculated to be 53.81%. In comparison to the current standards set by committee, the peer group set higher standards. They also showed an unexpected correspondence between percentage correct required for passing and the percentage actually passing. Comparison models suggested by Hofstee and Beuk, when applied to the survey data, indicated that 61-63% of the examinees should pass, with scores of 64-67% required to pass. (MGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED285914

Congruence of Standard Setting Methods for a Nursing Certification Examination

Lawrence J. Fabrey and Mark R. Raymond

Paper presented to the National Courcil on Measurement in Education
Annual Meeting, April 1987, Washington, DC.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

L.J. Fabrey

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM 870 522

Congruence of Standard Setting Methods for a Nursing Certification Examination

Lawrence J. Fabrey and Mark R. Raymond

Background

Certification agencies have increasingly adopted absolute standard setting procedures in favor of relative ones. In addition, methods for reaching a compromise between absolute and relative procedures have been proposed (DeGruijter, 1985). After classifying standard setting methods according to a judgmental-empirical continuum, Berk (1986) provided an evaluation of the technical adequacy and practicability of each. In selecting a method, Berk recommended the use of some form of judgmental analysis (for political reasons), and use of a conceptually and computationally simple technique (for the sake of credibility).

In describing judgmental methods, Livingston and Zieky (1982) state that standard setting judgments should be made in a way that accounts for the purpose of the test, by qualified persons, for whom the judgments have meaning. Berk (1986) suggested that standard setting issues for educational certification test specialists and licensure/certification boards are similar, except that with the latter group, sampling judges from a variety of populations is not necessary. Traditionally, official boards have been responsible for certification or licensure standard setting, but arguments could be made for involving other groups in the process. For example, if a certification program is intended for professional recognition, a peer group of examinees and/or certificants might be most appropriate for standard setting. On the other hand, if a program is primarily intended to protect the public, perhaps the public should help determine the requisite level of knowledge and skill.

The American Nurses' Association (ANA) certification program is intended primarily for professional recognition beyond initial licensure. The purpose of this study was to investigate the possible outcomes of asking a representative peer group of recent certificants to determine the examination passing score. Specifically, this study was designed to assess the degree of discrepancy between absolute, relative and compromise standards that would be set by certificants, and the extent to which application of the various standard setting models would approximate the actual passing point that had been set by committee for a recent examination.

Methods

A one-page survey was mailed in May 1985 to a random sample of 200 recently certified nurses. All respondents had met the same eligibility requirements to sit for the examination, had taken the same examination, and had exceeded the identical standard. The nurses had taken the examination the previous October and had received their score reports approximately four months preceding the survey.

The survey was designed to elicit perceptions regarding at what point a

---

Paper presented to the National Council on Measurement in Education Annual Meeting, April 1987, Washington, DC.

Authors' address: American Nurses' Association, 2420 Pershing Road, Kansas City, Missouri, 64108.

standard should be set by relative or absolute methods. Specifically, three questions were asked:

1) What percentage of examinees should pass the national certifying examination?
2) What percentage of questions should an examinee answer correctly in order to be certified?
3) Given the hypothetical distribution of scores on a 75-item test shown on the survey, what score should be achieved in order to be certified?

Selected background information was provided for each question. For example, for the first question, respondents were asked to bear in mind that criteria pertaining to eligibility (e.g. licensure, current practice) had been met. In addition, respondents were informed that "there is no correct answer; we are seeking your opinion about the proportion of your colleagues that should pass the examination."

For question two, respondents were asked to "bear in mind that it would be virtually impossible to get 100 percent correct, and that one would expect to get 25 percent correct by random guessing." They were to assume that the test questions are relevant and of varied difficulty levels. While the percentage correct is a somewhat crude statistic that may tend to perpetuate stereotypical standards, it was expected to be more understandable to the respondents, particularly since percent correct scores were among those provided on the score reports that respondents had received several months previously.

For question three, the distribution of scores shown was comparable to that of a recent candidate group in the certification area being surveyed; the shape was identical but the number of examinees and raw score values were changed. In addition to presenting the distribution of scores, the minimum, maximum, modal and mean scores were noted, to help ensure that respondents would understand the table.

Results

Eight weeks following the mailing, 98 usable responses were returned on the postage paid cards included in the mailing. Because the respondents were assured of complete anonymity, no follow-up was attempted.

Eight respondents chose not to answer the first question dealing with the percent that should pass. Most of these individuals indicated that anyone who can achieve a certain score should pass. The mean response of the 90 respondents to question one was 70.36 percent passing, and the standard deviation was 16.43 (standard error = 1.66). Responses ranged from 25 to 100 percent, in a negatively skewed distribution peaked with 14 responses at 75 percent, and 15 responses at 80 percent.

The mean percent correct value in response to the second question was 71.28, and the standard deviation was 9.72 (standard error = 0.98). Responses ranged from 40 to 95 percent correct, and the modal response was 75 ($n = 29$). While a negative correlation between percent passing and percent correct would be expected, a positive correlation of .21 ($p = .04$) was found.

While the first two questions encouraged the respondents to think in relative and absolute terms, respectively, the third question presented a distribution of

scores and provided no specific guidance. The mean response to question three (based on n = 96) was 53.81 (or 71.75 percent of the items correct), and the standard deviation was 7.01 (standard error = 0.71). Responses ranged from 22 to 70, and the distribution peaked at 55 (n = 16) and 56 (n = 15). The correlation between the percent correct and number correct responses was .49 ($p < .001$), while the correlation between percent passing and number correct was -.06 (n.s.).

The compromise models suggested by Hofstee (1983) and Beuk (1984) were applied to the data from the survey. With the Beuk model, the preferred combinations of *c* (cutoff score) and *f* (passing percentage) were drawn directly from responses to the first two questions on the survey. With the Hofstee method, judges are asked to provide the minimum and maximum acceptable cutoff scores, and the minimum and maximum acceptable percentage of failers. Since the survey in this study asked for the preferred cutoff and passing rate, and not acceptable ranges, the ranges were fabricated in three different ways: by using extreme values, trimmed extremes, and deviations from the mean preferred values derived from the first two questions on the survey.

The passing point resulting from application of the Beuk model was equivalent to 66 percent correct. Passing points resulting from the three modifications of the Hofstee method were: 67 percent correct using extreme values, 66 percent correct using trimmed extremes, and 64 percent correct using deviations from mean values.

Finally, the preceding results were compared to the actual standard. The standard setting committee had used a modified Angoff technique to set a passing score equivalent to 56 percent correct. For the 1984 test administration, this standard resulted in a passing rate of 87.96 percent.

Table 1
Comparison of Standard Setting Methods

Basis for pass point	Percent Correct	Passing Rate
Actual standard applied in 1984		
Modified Angoff	56	88
Survey results		
What percentage should pass?	*64*	70
What percent correct is enough?	71	*46*
Given data, what is the pass point?	72	*42*
Hofstee method (using extremes)	*67*	*61*
Beuk method	*66*	*63*

Note: Italicized numbers are estimated using 1984 score distribution (e.g., a passing point equivalent to 71 percent correct would have resulted in passing approximately 46 percent of the examinees).

Table 1 shows that applying the results of the survey (using any of the methods) to the 1984 score distribution would have resulted in a higher standard, and consequently, a lower passing rate. Assuring that 70.36 percent would pass, as indicated by question one, would have set a standard near 64 percent correct. On

the other hand, setting the standard at 71.28 percent correct (question two) would have resulted in a passing rate near 46 percent. Applying the mean number correct (53.81) for the third question to the 1984 distribution would have resulted in a passing rate near 42 percent. Finally, applying the modified compromise model results would have provided pass rates between 61 and 70 percent. Regardless of method, the consensus of judges responding to this survey was considerably more harsh than that of the standard setting committee.

Discussion

The major findings of this study include:

1) the relatively higher standard that would have been set by a peer group compared to that set by a committee,
2) an unexpected correspondence between the respondents' judgments regarding percent correct and percent passing, and
3) documentation of an application of compromise methods to achieve a compromise among certificants.

With actual data, as the percent correct is increased, more examinees fail to exceed the standard. Surprisingly, while a negative correlation would be expected between questions one and two, a positive one was found. It could be that many respondents were considering an ideal situation in which many candidates pass and also achieve high scores. Anecdotically, it is known that standard setting committees have similar illusions. However, alternative explanations for the inconsistency may be the lack of knowledge of the respondents regarding the usual relationship between the two variables, or misperceptions regarding the overall difficulty of the examination.

The correlation of .49 between percent correct and number correct (questions 2 and 3) was not surprising. This may be an indication that when a complete data set is displayed, respondents display the inclination to apply absolute standards. This interpretation would be consistent with the refusal of eight respondents to supply a relative standard.

Several factors could have contributed to the harshness of the respondents and have implications for the limitations of this study. The tendency of raters to be harsh while setting absolute standards is not a new phenomenon (see, for example, Schoon, Guillion and Ferrara, 1979). This tendency may have been reinforced because the individual respondent had already exceeded the standard. In fact, a useful replication of this study might include a mechanism for identifying the respondent's test score. It could be that respondents were applying a standard in some way relative to their own performance, for example, just below their score. The absolute standard set by the survey respondents was approximately equal to the mean test performance.

It could also be that the relatively low response rate provided a biased sample. The survey was short, but required the respondent to use analytical skills, which may have discouraged a portion of the sample. Identification of individual test scores or other characteristics of the respondents could provide an indication of whether or not a response bias existed. While the response rate was disappointing, those who did respond may have more closely matched Livingston and Zieky's (1982) criteria for judges; that is, the judgments of respondents evidently had greater personal meaning than the judgments of the non-respondents.

Another potential source of response bias relates to the timing of the survey.

As discussed previously, the perceptions of the difficulty of the examination may not have been accurate; the respondents took the examination approximately eight months previously, and received their results approximately five months before receiving the survey. For various reasons, it seemed inappropriate to distribute the survey at any time other than well after examinees had received their score reports. Future studies could be designed to seek judgments at other times, such as before or immediately following test administration.

Finally, the actual standard (equivalent to 56 percent correct) is relatively lower than that applied to other ANA certification examinations, and lower than other certification examinations, as well. It could be that judgments of certificants in other areas would differ. Future research may be directed toward addressing some of these factors, which may further the generalizability of the results.

It could be argued that delegating standard setting to examinees or certificants is an approach consistent with the purpose of some certification programs because it accounts for the purpose of the test, that is, professional recognition. Considerable risk could be involved if such a policy were implemented. However, examinee judgments regarding standards could be useful for setting, adjusting, and defending standards without drastically altering the traditional approach to identifying a passing point. After data have been collected from a representative candidate group, official boards could find the data useful in identifying a standard, or validating an existing standard. Considerations such as the timing of the data collection and sampling procedures may be useful in formulating future research.

References

Berk, R.A. (1986) A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research, 56(1)*, p.137-172.

Beuk, C.H. (1984) A method for reaching a compromise between absolute and relative standards in examinations. *Journal of Educational Measurement, 21(2)*, p. 147-152.

DeGruijter, D.N.M. (1985) Compromise models for establishing examination standards. *Journal of Educational Measurement, 22(4)*, p. 263-269.

Hofstee, W.K.B. (1983) The case for compromise in educational selection and grading. In S.B. Anderson & Helmick, J.S.D. (eds.), *On Educational Testing*. San Francisco: Jossey-Bass.

Livingston, S.A. and Zieky, M.J. (1982) *Passing Scores*. Educational Testing Service.

Schoon, C.G., Guillion, C.M. and Ferrara, P. (1979) Bayesian statistics, credentialing examinations, and the determination of passing points. *Evaluation and the Health Professions, 2*, p.181-201.